

ANNA WAŁEK
Politechnika Gdańska

Zarządzanie danymi badawczymi i ich udostępnianie – nowe zadania bibliotek akademickich na przykładzie Biblioteki Politechniki Gdańskiej

Streszczenie. Biblioteki akademickie zawsze były pionierami, jeśli chodzi o wdrażanie zasad otwartości, tworzenie narzędzi w postaci platform i repozytoriów, a także prowadzenie szkoleń dla kadry naukowej. Również w wypadku coraz intensywniejszego wdrażania polityki otwartego dostępu, w tym do danych badawczych, rola bibliotek akademickich i ich pracowników staje się kluczowa. Powstają nowe specjalizacje, takie jak bibliotekarz danych (*data librarian*) i *data steward*. Pracownicy tych specjalizacji nie tylko odpowiadają za szkolenia dla kadry naukowej z zagadnień dotyczących otwartych danych badawczych, ale oferują również bezpośrednią pomoc, świadczoną na rzecz zespołów badawczych w zakresie zarządzania danymi badawczymi i ich udostępniania.

Usługi bibliotek świadczone naukowcom i instytucjom badawczym w celu wspomagania ich w procesie zarządzania danymi badawczymi ewoluują. Nowe obowiązki bibliotek różnią się od dotychczasowych zadań. Uczestnicząc w zarządzaniu danymi, bibliotekarze stają się częścią procesu badawczego od samego początku cyklu życia danych. Ich wiedza i zaangażowanie są oczekiwane na różnych etapach, od tworzenia planów zarządzania danymi poprzez zbieranie, opisywanie, przechowywanie i udostępnianie danych. Dodatkowo bibliotekarze będą również odpowiedzialni za zapewnienie właściwego wykorzystania i cytowania zbiorów danych przygotowanych przez zespoły badawcze.

Największym wyzwaniem dla bibliotekarza przyszłości będzie zrównoważenie zapotrzebowania na zarządzanie danymi badawczymi ze wszystkimi innymi zadaniami, za które bibliotekarze są odpowiedzialni. Zwiększa się liczba zasobów elektronicznych, a drukowane książki nie znikną z większości bibliotecznych półek, więc bibliotekarze będą zmuszeni łączyć tradycyjne zadania z obowiązkami związanymi z szybko zmieniającą się technologią.

Celem artykułu jest przedstawienie działalności Centrum Kompetencji Otwartej Nauki w Bibliotece Politechniki Gdańskiej – pierwszej w Polsce jednostki aktywnie wspierającej proces zarządzania danymi na uczelni, a także świadczącej usługi doradcze i szkoleniowe na rzecz środowiska naukowego i bibliotekarskiego. Ma to na

celu przedstawienie nowej formy organizacyjnej i funkcjonalnej w bibliotekach akademickich.

Zastosowano metodę badań jakościowych – studium przypadku, która umożliwia poznanie i opisanie przypadków skrajnych, odstających od innych, znanych. Ma to na celu poszerzenie wiedzy na temat opisywanego zjawiska, a także dostarczenie danych porównawczych do dalszych badań nad nim.

Słowa kluczowe: otwarte dane badawcze, zarządzanie danymi badawczymi, biblioteki naukowe, otwarta nauka, MOST Wiedzy, MOST Danych, *data steward*, *data librarian*, zawód bibliotekarza.

Wstęp

Dane badawcze (*Research Data*) to wszystkie dane, które powstają w wyniku prac naukowo-badawczych, niezależnie od tego, czy mamy do czynienia z naukami medycznymi, ścisłymi, przyrodniczymi, społecznymi, czy humanistycznymi. Danymi mogą być m.in. wyniki analiz, pomiarów, ankiet, a także źródła historyczne, dzienniki laboratoryjne, obrazy i próbki.

Znaczenie udostępniania danych badawczych jest tematem szeroko dyskutowanym w środowisku naukowym i coraz powszechniejszym wśród badaczy. Dzieje się tak z jednej strony za sprawą instytucji finansujących badania naukowe, które zobowiązują swoich grantobiorców do dzielenia się danymi (Komisja Europejska, instytucje zrzeszone w Science Europe – np. Narodowe Centrum Nauki), z drugiej zaś strony dzięki dobrym praktykom, które stanowią wzór i inspirację dla innych badaczy. Dodatkowo ilość wytwarzanych danych stale rośnie. Pojawiają się pytania dotyczące ich przechowywania, zabezpieczania i udostępniania z wykorzystaniem odpowiednio przygotowanych do tego narzędzi.

Instytucje finansujące badania naukowe podkreślają, że znaczne fundusze przeznaczane na granty badawcze w dużej mierze się marnują. Dane, będące ich wynikami, nie są odpowiednio przechowywane ani udostępniane. Również wyniki publikowane w czasopiśmie naukowych trudno w takim wypadku zweryfikować, a badania, które zostały zakończone wraz z projektem, są niemożliwe do odtworzenia, kontynuowania i wykorzystania. Taka sytuacja powoduje np. to, że w wielu ośrodkach badawczych prowadzone są równoległe te same badania bądź ponownie te, które już zostały kiedyś przeprowadzone, nawet jeśli w przeszłości nie przyniosły oczekiwanych wyników lub można byłoby je kontynuować na podstawie zebranych i udostępnionych danych.

Przykładem działania promującego otwarty dostęp do wyników badań jest polityka Narodowego Centrum Nauki (NCN). Polityka Open Access NCN została wprowadzona 31 maja 2020 roku, a obowiązuje od konkursów ogłoszonych 15 czerwca 2020 roku. NCN promuje również otwarte udostępnianie danych badawczych. Od 2019 roku do wniosków grantowych składanych w konkursach NCN należy załączyć plan zarządzania danymi (PZD; *Data Management Plan* – DMP), który prezentuje m.in. to, jakie dane badawcze będą wykorzystywane, wytwarzane i udostępniane w ramach projektu, a także jak będą one przechowywane¹.

Zarówno środowiska naukowe, społeczeństwo, jak i biznes potrzebują dostępu do wysokiej jakości danych. Środowiska naukowe opierają swoje badania w dużej mierze na już zbadanych zjawiskach i wcześniejszych eksperymentach naukowych. W przypadku dostępu do wyników badań wyłącznie w formie publikacji naukowych w zasadzie niemożliwe jest zweryfikowanie tychże wyników ani wykorzystanie ich w innym kontekście, niż wynika to z zamierzeń autorów publikacji. Społeczeństwo wykorzystuje dane badawcze wprost, np. stosując je w życiu codziennym, a także do różnego rodzaju projektów z obszaru tzw. nauki obywatelskiej, jak również pośrednio, korzystając z odkryć nauki i techniki, a jednocześnie chcąc zgłębić zasady działania urządzeń czy pewnych praw i procesów. Szczególne znaczenie ma to w edukacji. Zapotrzebowanie na dzielenie się danymi badawczymi szczególnie silnie podkreśla się od początku pandemii COVID-19. Miało bowiem ogromny wpływ na postęp nauki w zakresie badań medycznych i farmaceutycznych². Jednym z kluczowych odbiorców danych są szeroko rozumiane biznes i przemysł, które w dużej mierze opierają swoje projekty badawczo-rozwojowe na współpracy z jednostkami naukowymi. Dostęp do wysokiej jakości danych może w tym przypadku z jednej strony być impulsem do nawiązania współpracy na linii nauka – biznes, a z drugiej strony stanowić początek współpracy w zakresie komercyjnego generowania i przekazywania danych (np. na potrzeby uczenia sztucznej inteligencji)³.

¹ Narodowe Centrum Nauki, *Instrukcja – Polityka NCN dotycząca otwartego dostępu do publikacji*, <https://tinyurl.com/shhkv9ee> [dostęp: 24.08.2021].

² I. Le Guillou, *Covid-19: How unprecedented data sharing has led to faster-than-ever outbreak research*, „Horizon. The EU Research and Innovation Magazine”, 23.03.2020, <https://tinyurl.com/82y3kaat> [dostęp: 20.09.2021]; C. Rudow, E. Baldi, *A Year of Data Insights in the Time of COVID-19*, Open Data Watch, <https://tinyurl.com/2p8pptdx> [dostęp: 20.09.2021].

³ *List of datasets for machine-learning research*, Wikipedia, <https://tinyurl.com/2p8cchhk> [dostęp: 20.09.2021].

Aby tę jakość zagwarantować, stworzono zasady FAIR DATA, które zawierają szereg wytycznych opartych na czterech głównych przesłankach, które tworzą akronim FAIR. Są to *Findability* (możliwość odnalezienia), *Accessibility* (dostępność), *Interoperability* (interoperacyjność), *Reusability* (możliwość ponownego wykorzystania). Zostały one sformułowane w 2016 roku. Zasady FAIR opisują różne kwestie dotyczące współczesnych środowisk publikowania danych w odniesieniu do wspierania zarówno ręcznego, jak i automatycznego osadzania, eksploracji, udostępniania i ponownego wykorzystania. W zasadach używa się wyrażenia (*meta*)dane w przypadkach, w których zasada powinna być stosowana zarówno do metadanych, jak i do danych. Elementy zasad FAIR są ze sobą powiązane, ale jednocześnie niezależne i rozdzielne. Zasady określają cechy, które powinny wykazywać współczesne zasoby danych, narzędzia, słowniki i infrastruktury, aby pomóc w ich odkrywaniu i ponownym wykorzystywaniu przez osoby trzecie. Dzięki minimalnemu zdefiniowaniu każdej zasady przewodniej bariera wejścia dla producentów danych, wydawców i specjalistów od danych, którzy chcą, aby ich zbiory danych były FAIR, jest celowo utrzymywana na jak najniższym poziomie. Zasad można przestrzegać w dowolnej kombinacji i stopniowo, w miarę jak środowiska wydawnicze dostawców danych ewoluują do coraz większego stopnia tzw. FAIRyfikacji. Ponadto modułowość zasad i ich rozróżnienie na dane i metadane pozwalają na zastosowanie szczególnych okoliczności. Jednym z takich przykładów są bardzo wrażliwe lub identyfikowalne dane, w przypadku których publikacja bogatych metadanych, w celu ułatwienia wyszukiwania, w tym jasnych zasad dotyczących procesu uzyskiwania dostępu do danych, zapewnia wysoki stopień FAIRyfikacji nawet w przypadku braku publikacji samych danych⁴.

Używanie zasad FAIR jako wytycznych do wyszukiwania i odczytywania danych przez ludzi i maszyny ułatwia ich ponowne wykorzystanie. Właściwie gromadzone, archiwizowane i rozpowszechniane dane zwiększają również współpracę międzynarodową i interdyscyplinarną⁵.

Dane powinny spełniać zasady FAIR, ale nie jest to jednoznaczne z tym, że dostęp do wszystkich powinien być otwarty. Podstawowa zasada udostępniania danych badawczych mówi, że powinny być one „na tyle otwarte, na ile to możliwe, i na tyle zamknięte, na ile to konieczne”. Nie

⁴ M. Wilkinson, M. Dumontier, I. Aalbersberg et al., *The FAIR Guiding Principles for scientific data management and stewardship*, „Scientific Data” 2016, t. 3 [dostęp: 20.09.2021].

⁵ V. de Arruda Jorge, S. Albagli, *Research data sharing during the Zika virus public health emergency*, „Information Research” 2020, t. 25, nr 1, <http://informationr.net/ir/25-1/paper846.html> [dostęp: 22.09.2021].

każde bowiem dane mogą czy powinny być udostępniane, nie wszystkie też nadają się do udostępniania.

Zasady FAIR powinny być implementowane w repozytoriach i infrastrukturach, które powstają w celu gromadzenia i udostępniania danych badawczych i/lub ich metadanych. Rozwiązania techniczne, takie jak otwarte protokoły, a także stosowanie schematów metadanych dobranych według określonych standardów pozwalają na zwiększenie poziomu wyszukiwalności, dostępności, interoperacyjności i możliwości ponownego użycia danych.

Należy jednak podkreślić, że poziom wdrożenia i liczba infrastruktur (komercyjnych i open source) różnią się w zależności od kraju. Według EU Open Science Monitor dostępność repozytoriów danych, obok polityk agencji finansujących badania naukowe oraz stosunku naukowców do udostępniania danych, to jeden z trzech podstawowych wskaźników związanych z Open Research Data⁶. Niemcy, które są europejskim liderem pod względem liczby repozytoriów otwartych danych, zindeksowały w Re3Data – Rejestrze Repozytoriów Otwartych Danych – 419 takich usług. Porównanie tej liczby do zaledwie sześciu polskich repozytoriów, które są tam indeksowane, wskazuje na konieczność tworzenia nowych polskich serwisów lub szerszego wykorzystania już działających⁷. Zapewnienie polskim naukowcom przestrzeni do archiwizacji i rozpowszechniania ich danych badawczych jest priorytetem i podstawowym warunkiem, aby dane te stały się częścią światowej komunikacji naukowej. Co więcej, repozytoria te muszą zachować standardy jakości, aby móc podlegać indeksacji w takich bazach, jak Data Citation Index⁸ lub Google Dataset Search⁹, co jest bardzo ważne z punktu widzenia budowania uznania naukowego. Obecnie w Web of Science Data Citation Index zindeksowane jest tylko jedno repozytorium z Polski – Katalog Danych Badawczych na platformie MOST Wiedzy, tworzony przez Politechnikę Gdańską¹⁰.

Różne badania, w tym to przeprowadzone przez Remediosa Melero i Caroline Navarro-Molinę¹¹, pokazują, że naukowcy chcą, aby ich dane

⁶ *Open Science Monitor 2019. Updated Methodological Note*, 4.04.2019, <https://tinyurl.com/2p8tuu2p> [dostęp: 22.09.2021].

⁷ Na podstawie: re3data.org [dostęp: 20.09.2021].

⁸ *Data Citation Index. Connecting data to the research it informs*, <https://tinyurl.com/22yca57x> [dostęp: 20.09.2021].

⁹ *Dataset Search*, <https://datasetsearch.research.google.com/> [dostęp: 20.09.2021].

¹⁰ [Mostwiedzy.pl](https://mostwiedzy.pl).

¹¹ R. Melero, C. Navarro-Molina, *Researchers' attitudes and perceptions towards data sharing and data reuse in the field of food science and technology*, „*Learned Publishing*” 2020, t. 33, nr 2, s. 163–179.

były cytowane. Mimo to często nie posiadają odpowiedniej wiedzy i umiejętności, aby zapewnić dostęp do danych we właściwy sposób. Ponadto badanie prowadzone przez Djogo Sayogo i Theresę Pardo¹² wykazało, że najważniejszymi determinantami dla badaczy w zakresie publikowania danych są umiejętności zarządzania danymi i wsparcie organizacji.

MOST Wiedzy

Podążając za światowymi trendami i zgodnie z zasadami otwartej nauki, w 2017 roku Politechnika Gdańska (PG) uruchomiła platformę MOST Wiedzy (mostwiedzy.pl). Portal powstał w ramach projektu „Multidyscyplinarny Otwarty System Transferu Wiedzy – MOST Wiedzy” i był współfinansowany ze środków Europejskiego Funduszu Rozwoju Regionalnego w ramach Programu Operacyjnego Polska Cyfrowa na lata 2014–2020.

Głównym celem projektu MOST Wiedzy było wdrożenie platformy integrującej dane z wielu baz danych Politechniki Gdańskiej oraz potencjalnie innych uczelni i ośrodków badawczych. Bardzo ważną częścią systemu jest otwarte repozytorium publikacji (repozytorium instytucjonalne OA). Pracownicy Politechniki Gdańskiej oraz inne osoby posiadające profil naukowy (możliwy do założenia na podstawie konta Google lub ORCID) mogą deponować publikacje, które następnie, po przejściu procesu weryfikacji w Bibliotece PG, są udostępniane w otwartym dostępie. Zgromadzone dane są dostępne publicznie, spójne i gotowe do ponownego wykorzystania. Platforma MOST Wiedzy służy odbiorcom za pomocą rozbudowanego systemu e-usług. Oprócz wyszukiwania i przeglądania informacji system posiada wbudowaną inteligencję oraz pewne mechanizmy analizy kontekstu wyszukiwania, co jest rozwiązaniem nowatorskim w naszym kraju. Dzięki temu platforma jest swego rodzaju inteligentnym konsultantem, podpowiadającym treści na podstawie historii wyszukiwania. Multidyscyplinarne e-usługi zlokalizowane są w prywatnej chmurze PG z ogólnodostępnym uniwersalnym API. Wszystkie zasoby zgromadzone w uczelnianych bazach wiedzy zostały udostępnione w ramach otwartego dostępu, co wpisuje się w inicjatywę otwartej nauki i stanowiło jeden z filarów projektu. Bazy te obejmują m.in. informacje o pracownikach naukowo-badawczych,

¹² D.S. Sayogo, T.A. Pardo, *Exploring the determinants of scientific data sharing: Understanding the motivation to publish research data*, „Government Information Quarterly” 2013, t. 30, s. 19–31.

przygotowanych przez nich pracach naukowych (w tym pełne treści), realizowanych projektach i grantach, posiadanej infrastrukturze naukowej (takiej jak laboratoria, aparatura), zespołach badawczych, opracowanych rozwiązaniach innowacyjnych, patentach, dostępnych unikalnych kursach online oraz ofercie technologicznej. MOST Wiedzy, rozpoczęty jako usługa instytucjonalna, jest obecnie dostępny dla wszystkich, którzy chcieliby zwiększyć widoczność dorobku naukowego, a także zdeponować publikacje w repozytorium¹³.

Kolejnym krokiem w rozwoju infrastruktury niezbędnej dla otwartej nauki było utworzenie otwartego repozytorium danych badawczych, aby zapewnić ich widoczność, dostępność i możliwość ponownego wykorzystania.

W 2018 roku do realizacji projektu „MOST Danych” (MOST DANYCH. Multidyscyplinarny Otwarty System Transferu Wiedzy. Etap II – Open Research Data) przystąpiły trzy największe pomorskie uczelnie: Politechnika Gdańska, Gdański Uniwersytet Medyczny i Uniwersytet Gdański. Projekt również jest współfinansowany ze środków Europejskiego Funduszu Rozwoju Regionalnego w ramach Programu Operacyjnego Polska Cyfrowa 2014–2020. Głównym jego celem jest zaprojektowanie i zbudowanie repozytorium danych, umożliwiającego zbieranie, wyszukiwanie, analizowanie i udostępnianie otwartych danych badawczych z uczestniczących instytucji. Repozytorium zostało uruchomione w kwietniu 2020 roku jako kolejna część platformy mostwiedzy.pl. Ponadto projekt obejmuje stworzenie Wirtualnego Mikroskopu, narzędzia wykorzystującego dane badawcze pochodzące z obrazowania medycznego¹⁴, bazy polityk Open Access polskich wydawców¹⁵, usługi publikowania czasopism Open Access oraz platformy do zarządzania konferencjami.

¹³ A. Wałek, P. Lubomski, *The Bridge to Knowledge – Open Access to Scientific Research Results on Multidisciplinary Open System Transferring Knowledge Platform*, „TASK Quarterly” 2017, t. 21, nr 4, s. 333–342.

¹⁴ Wirtualny Mikroskop – Cyfrowy Atlas Tkanek i Komórek to narzędzie umożliwiające przeglądanie w przeznaczonym do tego interfejsie danych z obrazowania medycznego, które pochodzą z komórek i tkanek różnych ludzkich narządów. Narzędzie to może służyć przede wszystkim w edukacji oraz kształceniu na poziomie wyższym, zawiera bowiem oprócz samych obrazów również opisy metadane obiektów, przygotowane przez patomorfologów, zawierające m.in. informacje o jednostkach chorobowych, widocznych na obrazie zmianach i sposobie barwienia oraz digitalizacji tkanek i komórek. Zob. <https://wirtualnymikroskop.mostwiedzy.pl/>.

¹⁵ Zob. K. Kokot, A. Szymik, *Zadania i kompetencje bibliotekarzy w zespole ds. polityk wydawniczych polskich czasopism naukowych*, „Biuletyn EBIB” 2021, nr 198, <http://ebib-boys.pl/index.php/ebib/article/view/738> [dostęp: 17.09.2021].

Twórcy projektu mieli świadomość, że budowanie wyłącznie infrastruktury technicznej bez wystarczającego wsparcia naukowców nie będzie efektywne i skuteczne. Dlatego od samego początku istotnym elementem projektu jest wsparcie merytoryczne oferowane pracownikom naukowym przez Centrum Kompetencji Otwartej Nauki (CKON), utworzone w Bibliotece Politechniki Gdańskiej jesienią 2018 roku.

Centrum Kompetencji Otwartej Nauki

Pomysł utworzenia Centrum Kompetencji zrodził się z globalnej tendencji zarządzania danymi oraz doświadczeń z poprzedniego projektu MOST Wiedzy. Okazało się bowiem, że wiedza badaczy na temat otwartej nauki jest niepełna i mają oni wiele pytań i obaw co do udostępniania wyników swoich badań w sposób otwarty. Są to wątpliwości natury zarówno prawnej (kwestie własności, licencji), jak i organizacyjnej (m.in. jakie dane, kiedy i gdzie należy udostępniać). Aby rozwiązać te praktyczne trudności, z jakimi borykają się badacze, w Bibliotece PG zorganizowano zespół, w skład którego wchodzi specjalista od danych (*data steward*) i bibliotekarze (*data librarian*).

Specjalizacja bibliotekarza danych nie została w pełni zdefiniowana. Według „Job Description Marketplace” organizacji LIBER bibliotekarz danych jest ekspertem w zakresie zarządzania danymi badawczymi, ich opisu, archiwizacji i rozpowszechniania. Musi on rozwinąć wiedzę techniczną dotyczącą praktycznych rozwiązań zarządzania danymi, archiwizacji i rozpowszechniania oraz eksploracji i wizualizacji danych. Bibliotekarze danych pomagają przede wszystkim projektować usługi dla naukowców i jednostek badawczych w zakresie zarządzania danymi. Wykorzystują ponadto swoją wiedzę, aby szkolić i wspierać badaczy w zakresie wszystkich aspektów związanych z zarządzaniem danymi, od definiowania potrzeb do proponowania praktycznych rozwiązań¹⁶.

Data steward to specjalizacja, a może nawet profesja nieco inna niż bibliotekarz danych. Jest to zjawisko stosunkowo nowe w polskim środowisku akademickim, ale również na świecie nadal toczą się dyskusje na temat profesjonalizacji tego zawodu. *Data steward* jest bezpośrednio odpowiedzialny za zarządzanie danymi i wspiera procesy cyklu życia danych. Jego rolą jest zasadniczo bezpośrednio wspieranie społeczności użytkowników. Odpowiada również za zbieranie, zestawianie i ocenę

¹⁶ LIBER Europe, *Data librarian, expert on research data management, description, archiving and dissemination* (2018), <https://tinyurl.com/2p8v2j4h> [dostęp: 13.10.2021].

problemów związanych z danymi. Zazwyczaj jest przydzielany według obszarów tematycznych, funkcji lub procesów biznesowych. Ponieważ jest on odpowiedzialny za jakość danych, musi móc bezpośrednio zaangażować się w proces powstawania danych i ich procesowania. Dlatego też jednym z kluczowych warunków zdefiniowania jego roli jest znajomość dyscyplin i procesów badawczych. Zasadniczo *data steward* jest kanałem komunikacji problemów związanych z cyklem życia danych – tworzeniem, modyfikacją, udostępnianiem, ponownym wykorzystaniem, przechowywaniem i tworzeniem kopii zapasowych danych. Jeśli pojawią się jakiegokolwiek problemy dotyczące zgodności danych ze zdefiniowanymi zasadami w okresie ich istnienia, jego obowiązkiem jest ich rozwiązanie¹⁷.

Biorąc pod uwagę powyższe kwestie, należy podkreślić, że jednym z głównych wyzwań Centrum Kompetencji było utworzenie zespołu profesjonalistów, którzy mogą zapewnić wsparcie społeczności akademickiej we wszystkich kwestiach związanych z Otwartą Nauką, zwłaszcza z otwartymi danymi badawczymi. Obecnie kadry naukowej, doktorantom i studentom oferowane są różne rodzaje szkoleń, takie jak szkolenia dedykowane (np. dotyczące dyscypliny naukowej), konsultacje bezpośrednio czy webinaria online z podziałem na bloki tematyczne (otwieranie danych badawczych, plany zarządzania danymi, wsparcie prawne – licencjonowanie danych, ich ochrona i ponowne wykorzystanie), szkolenia na temat zasad korzystania z repozytorium danych (deponowanie zestawów danych), a także zagadnień takich jak dane i metadane w kontekście zasad FAIR, planowanie i implikacje tzw. Planu S¹⁸ itp.

Kolejnym głównym wyzwaniem stojącym przed Centrum Kompetencji było uwzględnienie różnic między dyscyplinami naukowymi i ich różnymi praktykami komunikacji naukowej w zakresie udostępniania wyników naukowych.

Aby usługi i narzędzia tworzone w ramach projektu spełniały oczekiwania środowisk, którym mają służyć, konieczne było włączenie do prac nad jego realizacją zespołów badawczych ze wszystkich trzech uczelni.

Współpraca z naukowcami z różnych dyscyplin zaowocowała opracowaniem ogólnego standardu metadanych wykorzystywanych przez repozytorium Katalog Danych Badawczych (MOST Danych), a także

¹⁷ A. Wałek, *Data librarian and data steward – new tasks and responsibilities of academic libraries in the context of Open Research Data implementation in Poland*, „Przegląd Biblioteczny” 2019, z. 4, s. 497–512.

¹⁸ S. Coalition, *Plan S. Making full and immediate Open Access a reality*, <https://www.coalition-s.org/> [dostęp: 17.09.2021].

przygotowaniem dostosowanych do indywidualnych potrzeb szkoleń dla poszczególnych obszarów naukowych. Konieczne są bowiem różne podejścia, inne problemy pojawiają się, gdy mamy do czynienia m.in. z danymi humanistycznymi, inne w wypadku danych medycznych.

W realizację projektu zaangażowanych było ok. 60 naukowców reprezentujących różne dziedziny i dyscypliny naukowe. Ich zadaniem była inwentaryzacja ilości i typów danych występujących w ich macierzystych jednostkach, przygotowanie próbek danych wraz z opisem na potrzeby testowania repozytorium, a także udostępnienie danych. Do końca trwania okresu realizacji projektu (grudzień 2021 roku) wszystkie zespoły mają łącznie zdeponować ok. 30 tys. zestawów danych badawczych.

W zasobach repozytorium można znaleźć m.in. dane obrazowe, wyniki pomiarów i eksperymentów, dane medyczne, środowiskowe i biologiczne, a także z zakresu nauk społecznych, humanistycznych i nauk o sztuce.

Repozytorium danych

Udostępnianie danych w repozytorium danych badawczych na platformie MOST Wiedzy jest intuicyjne i proste dzięki odpowiednio zaprojektowanemu interfejsowi, wyposażonemu we wszystkie niezbędne funkcjonalności.

Korzystając z dedykowanego szablonu, badacze, którzy wcześniej założyli profile na platformie MOST Wiedzy, mogą przysyłać zestawy danych do repozytorium danych. Zestawy te muszą być opisane metadanymi w odpowiedni sposób, umożliwiając ich wyszukiwanie i ponowne wykorzystanie – a więc spełniającymi również zasady FAIR. Opis metadany, oparty na standardzie DDI¹⁹, składa się z jedenastu pól, które należy wypełnić w języku angielskim, niezależnie od tego, w jakim języku są same dane. Jest to warunek konieczny, jeśli repozytorium ma być indeksowane, zwłaszcza przez serwisy takie jak Data Citation Index (Web of Science). W procesie deponowania wyróżniamy cztery etapy weryfikacji zbioru danych: 1) w przygotowaniu, 2) w trakcie weryfikacji, 3) potrzeba rewizji, 4) przyjęty.

Po zdeponowaniu zbioru danych przez autora (deponenta) i opisanu go metadanymi zostaje on przesłany do weryfikacji do Centrum Kompetencji. Pracownicy Centrum weryfikują poprawność wyświetlenia pliku,

¹⁹ J. Ryssevik, *The Data Documentation Initiative (DDI) metadata specification*, Metanet, <https://tinyurl.com/2p9c8tbu> [dostęp: 17.09.2021].

jego format, licencję oraz przygotowany opis metadanych. Jeśli którykolwiek z elementów wymaga poprawy, rekord zostaje odesłany do deponenta. Po dokonaniu przez niego korekty i ponownej, pozytywnej weryfikacji Centrum rekord zostaje udostępniony.

Może to nastąpić od razu, po upływie embargo lub w określonym przez autora momencie. MOST, podobnie jak inne najlepsze otwarte repozytoria danych badawczych i usługi, m.in. Figshare, zapewnia również opcję ograniczonego dostępu do zbiorów danych. Na przykład w przypadku projektów rządowych lub procesów publikacji objętych embargiem badacze mogą wygenerować prywatny link, który można udostępnić tylko członkom zespołu lub wydawcom. Tego rodzaju zbiory danych muszą być przygotowywane etapowo przez cały okres obowiązywania embarga.

Kolejną bardzo ciekawą i przydatną funkcjonalnością dostępną w repozytorium jest możliwość deklarowania indywidualnej licencji. Badacz umieszcza wszystkie wymagania swojej licencji w pliku tekstowym i przesyła na etapie przygotowania. Po zaakceptowaniu zbioru danych licencja jest dostępna wraz ze wszystkimi informacjami o metadanych.

Plany zarządzania danymi

Poza wspieraniem procesu deponowania danych w repozytorium oraz organizacji szkoleń i konsultacji z zakresu ich udostępniania pracownicy Centrum Kompetencji są intensywnie zaangażowani również w pomoc naukowcom przygotowującym wnioski grantowe na potrzeby projektów Narodowego Centrum Nauki, Komisji Europejskiej (Horyzont 2020) oraz innych agencji fundingowych.

Zarządzanie danymi badawczymi nie jest dla naukowców niczym nowym, ponieważ jest ono niezbędne dla zachowania standardów prowadzonych badań i zapewnienia wiarygodności wyników. Nowość może stanowić obowiązek zaplanowania i przedstawienia wszystkich aspektów związanych z danymi badawczymi już na etapie składania wniosku.

Niektóre instytucje finansujące określiły wymagania, które powinny spełniać plany zarządzania danymi wytworzonymi podczas projektu badawczego. Plany te pokazują, w jaki sposób dane badawcze będą gromadzone, organizowane, zarządzane i przechowywane w trakcie projektu i po jego zakończeniu. To, jak taki plan wygląda, zależy od specyfiki i warunków projektu. Ogólnie rzecz biorąc, plany te wymagają opisu danych, które będą wytwarzane lub wykorzystywane, formatów, standardów metadanych, które będą wykorzystywane do przechowywania i organizowania danych, sposobu i miejsca przechowywania danych oraz dostępności.

Na plan zarządzania danymi w konkursach NCN składa się sześć punktów, które dotyczą poszczególnych etapów gromadzenia, przechowywania i udostępniania danych. Są to:

- 1) opis danych oraz pozyskiwanie lub ponowne wykorzystanie dostępnych danych,
- 2) dokumentacja i jakość danych,
- 3) przechowywanie i tworzenie kopii zapasowych podczas badań,
- 4) wymogi prawne, kodeksy postępowania,
- 5) udostępnianie i długotrwałe przechowywanie danych,
- 6) zadania związane z zarządzaniem danymi oraz zasoby²⁰.

Pomimo pewnej praktyki w zarządzaniu danymi badawczymi naukowcy często mają problemy z wypełnieniem poszczególnych rubryk planu. Problemy te można podzielić na dwie grupy:

- 1) brak wiedzy na temat bezpiecznego przechowywania danych badawczych, zasad ich udostępniania, otwartych licencji i zasad własności intelektualnej oraz metadanych i standardów;
- 2) niechęć do udostępniania danych badawczych, traktowanie planu jako kolejnego elementu zbędnej „papierologii” i „biurokracji”, strach naukowców przed przywłaszczeniem lub niewłaściwym wykorzystaniem ich danych przez innych²¹.

Powyższe problemy objawiają się często ambiwalentnym stosunkiem naukowców do samego procesu tworzenia planu i dążeniem do zrzucenia z siebie tej odpowiedzialności. Pracownicy Centrum Kompetencji często słyszą od naukowców, że tworzenie danych ich nie dotyczy. Dopiero po jakimś czasie okazuje się, że gromadzą czy też wytwarzają dane, które należy uwzględnić w planie. Inną przyczyną niechęci wobec wypełniania PZD jest mylne utożsamianie wszystkich danych powstałych w trakcie trwania projektu z otwartymi danymi badawczymi. Wyobrażenie, że wszystkie dane będą musiały znaleźć się w otwartym dostępie, zniechęca do rzetelnego przygotowania planu. Dlatego tak ważne jest, żeby na wstępie dokładnie wyjaśnić, czym jest PZD, czemu służy i czego dotyczy. Inną kwestią, którą należy zaznaczyć już na samym początku, jest to, że to wnioskodawca powinien być autorem PZD, zatem za jego skonstruowanie odpowiada kierownik projektu lub członek zespołu badawczego. Niejednokrotnie zdarza się, że naukowcy dostarczają pracownikowi Centrum Kompetencji podstawowe informacje na temat swoich badań,

²⁰ Narodowe Centrum Nauki, op. cit.

²¹ P. Krajewski, M. Szufliata-Żurawska, *Plan zarządzania danymi – jak skutecznie współpracować z naukowcami? Doświadczenia Centrum Kompetencji Otwartej Nauki przy Bibliotece Politechniki Gdańskiej*, „Biuletyn EBIB” 2020, nr 5 (194), s. 1–9.

oczekując, że przygotuje on plan za nich. Jednakże plan powinien opracowywać specjalista w dziedzinie, której ten plan dotyczy. Tylko on wie, jak zapewnić jakość danych w trakcie trwania prac badawczych, jakie formaty plików będą gromadzone i gdzie będą przechowywane. Rolą bibliotekarza jest udzielenie wsparcia i ewentualne uzupełnienie planu o elementy, które dla naukowców mogą być niejasne, np. standard metadanych użyty do opisu danych.

Należy pamiętać, że wniosek konkursowy zawiera jedynie skrócony plan. Na zakończenie projektu konieczne będzie dołączenie szczegółowego raportu opisującego, w jaki sposób poszczególne zapisy z wniosku zostały zrealizowane. Z tego powodu ważna jest ewidencja wpływających do Centrum Kompetencji wniosków, która umożliwi monitorowanie terminów dla poszczególnych projektów oraz szybkie odnalezienie zatwierdzonego wcześniej planu. W celu usprawnienia tego procesu stworzono bazę, do której mają dostęp wszyscy członkowie zespołu CKON. Baza zawiera m.in. imię i nazwisko autora, jednostkę, edycję konkursu, w ramach którego powstał plan, jest tu również miejsce na dołączenie pliku z planem.

Działalność szkoleniowa

W lutym 2020 roku, po wprowadzeniu obowiązku przygotowania PZD przez wnioskodawców, NCN zorganizowało dwa szkolenia dla bibliotekarzy i pracowników administracji uczelni, wspierających naukowców w przygotowywaniu wniosków grantowych. Celem szkoleń było omówienie poszczególnych części PZD oraz zasad otwartości w dostępie do danych. Ekspertami szkolącymi byli pracownicy Biblioteki PG. W odpowiedzi na duże zapotrzebowanie na wiedzę w tym zakresie oraz ze względu na brak odpowiednio wyszkolonych kadr w polskich instytucjach naukowych Centrum Kompetencji zorganizowało serię szkoleń w Gdańsku, a następnie online, zarówno dla bibliotekarzy i pracowników administracji, jak i dla naukowców reprezentujących różne dyscypliny naukowe. Łącznie przez Centrum Kompetencji przeszkolonych zostało już kilkaset osób z bibliotek i jednostek administracji, a także naukowców z wielu instytucji i uczelni z całej Polski. Od 2017 roku Biblioteka Politechniki Gdańskiej organizuje również cykliczną międzynarodową Pomorską Konferencję Open Science²². W każdej edycji prezentowane są najnowsze badania oraz praktyki dotyczące zarządzania danymi badawczymi.

²² Strona konferencji: <https://pg.edu.pl/pkos/strona-glowna>.

Materiały z konferencji udostępniane są w sposób otwarty, w tym na kanale Biblioteki PG w serwisie YouTube²³.

Podsumowanie

Usługi bibliotek świadczone naukowcom i instytucjom badawczym w celu wspomagania ich w procesie zarządzania danymi badawczymi ewoluują. Nowe obowiązki bibliotek różnią się od dotychczasowych zadań. Biblioteki muszą dostosować się do nowych warunków, aby być na bieżąco z usługami świadczonymi w zakresie zarządzania danymi. Bibliotekarze stają się częścią procesu badawczego od samego początku cyklu życia danych. Ich wiedza i zaangażowanie są niezbędne już na wczesnym etapie tworzenia PZD, w trakcie zbierania, opisywania, przechowywania i „kuratorstwa” danych. Dodatkowo bibliotekarze będą również odpowiedzialni za zapewnienie właściwego wykorzystania i cytowania zbiorów danych przygotowanych przez zespoły badawcze. Nowe zadania związane z zarządzaniem danymi badawczymi stawiają przed bibliotekarzami różne wyzwania, którym będą musieli sprostać.

Ważnym zadaniem bibliotek akademickich i zatrudnianych w nich nowych specjalistów – bibliotekarzy danych – będzie ścisła współpraca z zespołami badawczymi i innymi specjalistami zajmującymi się danymi (np. *data steward*).

Wzorem niektórych uczelni w innych krajach, takich jak Cambridge czy Delft, na polskich uczelniach powinny powstać zespoły multidyscyplinarne. Obowiązkiem tychże zespołów będzie wspieranie badaczy w zadaniach związanych z zarządzaniem danymi. Rolą biblioteki akademickiej będzie w przyszłości koordynacja pracy takich wielozadaniowych zespołów odpowiedzialnych za tworzenie narzędzi repozytoriów. Rola ta będzie również obejmować rozwój instytucjonalnych usług zarządzania danymi badawczymi oraz świadczenie usług doradczych i szkoleniowych. Te nowe zadania i działania pociągną za sobą potrzebę współpracy z innymi zainteresowanymi stronami, aby zapewnić skuteczne wspieranie przyszłych badań.

Przykładem takiej działalności jest Centrum Kompetencji Otwartej Nauki w Bibliotece Politechniki Gdańskiej. Szerokie spektrum działań Centrum Kompetencji, takich jak wsparcie w procesie deponowania danych i tworzenia planów zarządzania danymi, działalność szkoleniowa i konsultacyjna, a także uczestnictwo w działaniach międzynarodowych

²³ Adres kanału: <https://tinyurl.com/375vatpt>.

organizacji i grup roboczych, sprawia, że jest ono prekursorem i liderem nowych rozwiązań i specjalizacji w polskich bibliotekach akademickich.

Bibliografia

- Arruda Jorge V. de, Albagli S., *Research data sharing during the Zika virus public health emergency*, „Information Research” 2020, t. 25, nr 1, <http://informationr.net/ir/25-1/paper846.html> [dostęp: 22.09.2021].
- Coalition S., *Plan S. Making full and immediate Open Access a reality*, <https://www.coalition-s.org/> [dostęp: 17.09.2021].
- Data Citation Index. Connecting data to the research it informs*, <https://tinyurl.com/22yca57x> [dostęp: 20.09.2021].
- Dataset Search*, <https://datasetsearch.research.google.com/> [dostęp: 20.09.2021].
- Kokot K., Szymik A., *Zadania i kompetencje bibliotekarzy w zespole ds. polityk wydawniczych polskich czasopism naukowych*, „Biuletyn EBIB” 2021, nr 198, <http://ebibojs.pl/index.php/ebib/article/view/738> [dostęp: 17.09.2021].
- Krajewski P., Szuflita-Zurawska M., *Plan zarządzania danymi – jak skutecznie współpracować z naukowcami? Doświadczenia Centrum Kompetencji Otwartej Nauki przy Bibliotece Politechniki Gdańskiej*, „Biuletyn EBIB” 2020, nr 5 (194), s. 1–9. DOI: 10.5281/zenodo.4281418.
- Le Guillou I., *Covid-19: How unprecedented data sharing has led to faster-than-ever outbreak research*, „Horizon. The EU Research and Innovation Magazine”, 23.03.2020, <https://tinyurl.com/82y3kaat> [dostęp: 20.09.2021].
- LIBER Europe, *Data librarian, expert on research data management, description, archiving and dissemination* (2018), <https://tinyurl.com/2p8v2j4h> [dostęp: 13.10.2021].
- List of datasets for machine-learning research*, Wikipedia, <https://tinyurl.com/2p8cchhk> [dostęp: 20.09.2021].
- Melero R., Navarro-Molina C., *Researchers’ attitudes and perceptions towards data sharing and data reuse in the field of food science and technology*, „Learned Publishing” 2020, t. 33, nr 2, s. 163–179. DOI: <https://doi.org/10.1002/leap.1287>.
- Mostwiedzy.pl [dostęp: 20.09.2021].
- Narodowe Centrum Nauki, *Instrukcja – Polityka NCN dotycząca otwartego dostępu do publikacji*, <https://tinyurl.com/shhkv9ee> [dostęp: 24.08.2021].
- Open Science Monitor 2019. Updated Methodological Note*, 4.04.2019, <https://tinyurl.com/2p8tuu2p> [dostęp: 22.09.2021].
- re3data.org [dostęp: 20.09.2021].
- Rudow C., Baldi E., *A Year of Data Insights in the Time of COVID-19*, Open Data Watch, <https://tinyurl.com/2p8pptdx> [dostęp: 20.09.2021].
- Ryssevik J., *The Data Documentation Initiative (DDI) metadata specification*, Metanet, <https://tinyurl.com/2p9c8tbu> [dostęp: 17.09.2021].
- Sayogo D.S., Pardo T.A., *Exploring the determinants of scientific data sharing: Understanding the motivation to publish research data*, „Government

- Information Quarterly” 2013, t. 30, s. 19–31. DOI: <https://doi.org/10.1016/j.giq.2012.06.011>.
- Walek A., *Data librarian and data steward – new tasks and responsibilities of academic libraries in the context of Open Research Data implementation in Poland*, „Przegląd Biblioteczny” 2019, z. 4, s. 497–512. DOI: 10.36702/pb.634.
- Walek A., Lubomski P., *The Bridge to Knowledge – Open Access to Scientific Research Results on Multidisciplinary Open System Transferring Knowledge Platform*, „TASK Quarterly” 2017, t. 21, nr 4, s. 333–342. DOI: <https://doi.org/10.17466/tq2017/21.4/e>.
- Wilkinson M., Dumontier M., Aalbersberg I. et al., *The FAIR Guiding Principles for scientific data management and stewardship*, „Scientific Data” 2016, t. 3. DOI: <https://doi.org/10.1038/sdata.2016.18>.
- Wirtualny Mikroskop – Cyfrowy Atlas Tkanek i Komórek, <https://wirtualnymikroskop.mostwiedzy.pl/>.

ANNA WAŁEK

Research data management and sharing – new assignments for academic libraries, with the example of the Gdańsk University of Technology Library

Abstract. Academic libraries have always pioneered in implementing the principles of openness, creating appropriate tools such as web platforms and institutional repositories, and providing training to academic and research staff. Correspondingly, the same pattern applies to the more and more intense implementation of the Open Access policy that also includes research data. In either assignments, the role of academic libraries and their staff is vital. Consequently, new specialisations emerge, such as the data librarian or data steward, who are responsible not only for providing support to the faculty through instruction and consultation in issues related to Open Research Data but also offer direct assistance for research teams in data management and data sharing. Library services that support researchers and research institutions across a wide spectrum of disciplines in managing research data are constantly evolving. As a result, new assignments for libraries tend to be different from those previously attached to them. Being part of the data management process in an institution, librarians are beginning to get seriously involved in the research process from the very onset of the data life cycle. Their expertise and involvement are expected at various stages, from creating Data Management Plans to retrieving, describing, storage, and providing access to data. In addition, librarians will also be responsible for providing the environment for appropriate use and citation of data sets prepared by research teams.

The biggest challenge for the future librarian will be to balance the demand for research data management with all other library tasks for which librarians are usually

responsible. The number of electronic resources is steadily growing, while traditional printed books will not disappear from the shelves. As a result, librarians will be necessitated to match traditional assignments with the new responsibilities related to the progress and advances in technology.

This article presents the activities of the Open Science Competence Center in the Library of the Gdańsk University of Technology – the first unit of this kind in Poland that actively supports the process of data management at the university and provides services in advice, assistance and training for the research and library communities. In this way, a new organisational and functional form of a library unit emerging in academic libraries is presented and discussed.

The qualitative research method was used – a case study, which makes it possible to identify and describe extreme cases, differing from others, already known. This, no doubt, expands our knowledge of the phenomenon described and provides additional comparative data for further studies.

Keywords: Open Research Data, research data management, research libraries, Open Science, MOST Wiedzy, MOST Danych, data steward, data librarian, librarian profession.

Tekst wpłynął do Redakcji 29 października 2021 roku.

